Kepler Science Operations Processes, Procedures, and Tools

Jennifer R. Hall[*a], Khadeejah Ibrahim[a], Todd C. Klaus[a], Miles T. Cote[b], Christopher Middour[a], Michael R. Haas [b], Jessie L. Dotson [b], Brett Stroozas[c], Michael Wu[d], Jeneen Sommers[a], Paresh Bhavsar[b]

[a]Orbital Sciences Corporation/NASA Ames Research Center, MS 244-30, Moffett Field, CA, USA 94035
[b]NASA Ames Research Center, MS 244-30, Moffett Field, CA, USA 94035
[c]Stroozas FlightOps, 420 Augustus Ct. Walnut Creek, CA, USA 94598
[d]Bastion Technologies, NASA Ames Research Center, MS 244-30, Moffett Field, CA, USA 94035

ABSTRACT

The Kepler Science Operations Center (SOC) is responsible for the configuration and management of the SOC Science Processing Pipeline, processing of the science data, distributing data and reports to the Science Office, exporting processed data for archiving to the Data Management Center at the Space Telescope Science Institute, and generation and management of the target and aperture definitions. We present an overview of the SOC procedures and workflows for the data the SOC manages and processes. There are several levels of reviews, approvals, and processing for the various types of data. We describe the process flow from data receipt through data processing and export, as well as the procedures in place for accomplishing the tasks. The tools used to accomplish the goals of *Kepler* science operations will be presented and discussed as well. These include command-line tools and graphical user interfaces, as well as commercial products. The tools provide a wide range of functionality for the SOC including pipeline operation, configuration management, and process workflow implementation. For a demonstration of the Kepler Science Operations Center's processes, procedures, and tools, we present the life of a quarter's worth of data, from target and aperture table generation through archiving the data collected with those tables.

Keywords: *Kepler*, science operations, operational processes, workflows, pipeline

1. INTRODUCTION

The *Kepler* spacecraft was launched into a 372.5-day earth-trailing, heliocentric orbit on March 7, 2009[1]. Approximately every 93 days, the spacecraft is rolled 90 degrees about its axis to keep the solar arrays illuminated and the focal-plane radiator away from the sun. *Kepler*'s photometer is an array of 42 CCDs with a total of 94.6 million pixels and a field of view covering 115.6 square degrees of the sky in the constellations Cygnus and Lyra. The primary goal of the *Kepler Mission* is to detect Earth-like planets transiting Sun-like stars by performing aperture photometry on the approximately 160,000 targets observed each quarter.

The Kepler Science Operations Center (SOC) is located at the NASA Ames Research Center in Moffett Field, California. The primary responsibilities of the SOC are to generate the science tables sent to the spacecraft, monitor the spacecraft health and performance, and process the science data returned from the spacecraft. This paper discusses the routine processing activities the SOC performs in section 2, below. In section 3, we discuss the processes, workflows, and procedures that define the SOC activities, with references to the operational activities introduced in section 2. In section 4, we present the tools employed at the SOC to perform routine activities. These include both graphical user interfaces to control the software and configuration management, and SOC-developed command-line tools. Commercially available software is used to manage the hardware and data products. The processes and tools employed enable the SOC to successfully perform science operations in support of the *Kepler* mission goals.

[*] jennifer.hall@nasa.gov

2. ROUTINE SCIENCE OPERATIONS ACTIVITIES

The timeline for the processing activities involved with a single quarter's worth of data is presented in figure 2. Each quarter, the SOC and Science Office (SO) work closely to generate the tables that define the pixel data to collect science on the spacecraft. A small subset of pixels is collected from the spacecraft during semi-weekly contacts to monitor the health and performance of the instrument. The table generation and semi-weekly processing activities are performed in a separate environment from the science processing activities, as described in section 2.1. Monthly contacts with the spacecraft return the science and engineering data from the spacecraft, which is then processed at the SOC. This monthly processing of the data provides further characterization of the instrument performance and preliminary results to support the quarterly processing of data. The products of the quarterly reprocessing of the data are reviewed by the Data Analysis Working Group (DAWG) at Ames, then delivered for archiving at the Data Management Center (DMC) at the Space Telescope Science Institute (STScI) in Baltimore, MD. The DAWG is a team of SO scientists and SOC developers who review the data processing parameters for optimization of the software performance and the data products delivered to the archive. The quarterly processing also identifies targets of interest for the Science Office to analyze for follow-up observations.

2.1 Hardware Resource Allocation

There are three processing clusters[2] in the SOC, as shown in Figure 1. In the *Flight Operations* (*FlightOps*) cluster we perform the flight critical functions: target and aperture table generation (or target management), Photometer Data Quality (PDQ) processing, and generation of the data compression tables. We have two *Science Processing* clusters - the *Monthly* and *Quarterly* clusters, SciProcM and SciProcQ, respectively. Processing of the science data is performed in parallel with the flight critical activities. In section 4 we discuss the user interface tools that the operator uses to process the data. In Figure 1, the relationships and roles of each cluster are shown. Each cluster has a Kepler Database (KDB), a relational database (DB), and a host of worker machines. The Kepler Database was developed at the SOC to store the pixel data and the products generated from processing the raw data[3]. The relational database stores supplemental data products, such as instrument models used in the data processing. In section 4.4, the tools used to manage and maintain the clusters are discussed. A snapshot of the FlightOps KDB is mounted on a regular basis as a secondary, read-only, DB in the SciProcQ cluster. This provides the Science Office with access to the processing results from the FlightOps cluster. A snapshot of the SciProcQ cluster is used in the SciProcM cluster, allowing for parallel processing of newly downlinked monthly data in the SciProcM cluster and processing of the previous quarter in the SciProcQ cluster. The SciProcM and SciProcQ Clusters usually operate with a different version of the software than the FlightOps Cluster. The FlightOps Cluster is "locked down" to ensure a stable code base for support of the flight critical functions at the SOC, such as generation of the target and aperture tables.

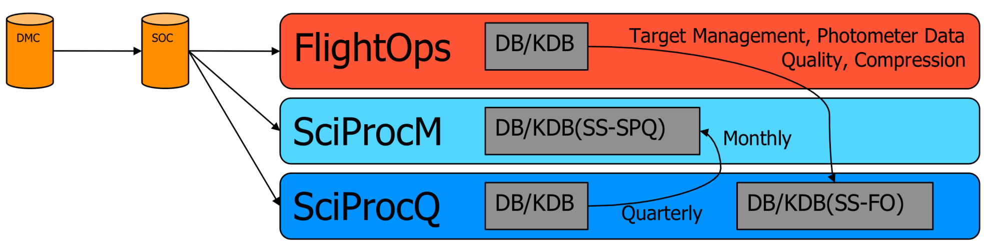

Figure 1. SOC Cluster Organization

2.2 Target Table Generation

The SOC works with the Science Office (SO) to specify the stellar targets to be uplinked to the spacecraft. Target management is one of the core activities, and consists of target selection, generation of target and aperture definitions, uplink of target/aperture definition tables to the spacecraft, and performance verification for those newly uplinked tables. A set of eight tables is generated for each quarter: six target tables (Long Cadence, Reference Pixel, Background, and three one-month Short Cadence tables), and two Aperture tables (target and background). All will be used during the entire quarter, with the Short Cadence tables swapped out on a monthly basis. The target selection process provides lists of prioritized targets of various categories that are finely balanced to fill the final tables. Although the planetary target list accounts for the vast majority of the science targets, a number of other lists are included to provide photometer performance information (e.g., stellar, and image artifact targets), while others perform different scientific duties (e.g., comparison, Guest Observer, and eclipsing binaries targets).

Because of storage and bandwidth constraints on the spacecraft, only ~5% of the total pixels collected are downlinked to Earth. The SOC is responsible for specifying which pixels to downlink, based on lists of stellar targets provided by the Science Office. See Figure 3 for the Target Management Activity workflow. The pixels of interest are specified via the target and aperture definition (TAD) tables[4] loaded onto the spacecraft for science data collection. These TAD tables are generated at the SOC during the Target Management activity (see Figure 2), which begins about six weeks prior to the quarterly roll that precedes the start of the quarter. The TAD tables are generated through a series of TAD pipeline runs in the Flight Operations data processing cluster. The activity produces target and aperture definition tables for the following types of science data: long cadence (LC), short cadence (SC), reference pixels (RP), background pixels, and aperture tables. Long cadence and reference pixel targets both have a 29.4-min sampling interval. The RP data is only extracted and saved every 48 LCs, about once per day. The short cadence data has a sampling interval of 58.8 sec. Background pixels are collected for use in calibration of the science data. The target management activity makes use of nearly every tool the SOC has developed, as well as open source software: the Pipeline Console, Target Management Console, and Mission Report Graphical User Interfaces; SOC-developed command-line tools; version control software; and an issue tracking tool. The way in which we use these tools will be detailed in sections 3 and 4 below. Once the TAD tables have been generated, they are approved by the Science Office, exported, and delivered to the Mission Operations Center (MOC) at the Laboratory for Atmospheric and Space Physics (LASP) in Boulder, CO and to the DMC. The DMC generates Pixel Mapping Reference Files (PMRFs) from the target/aperture tables and delivers them to the SOC. These files map the pixels to be collected on the spacecraft to their corresponding target and aperture definitions. Meanwhile, the MOC converts and formats the target/aperture tables for upload to the spacecraft.

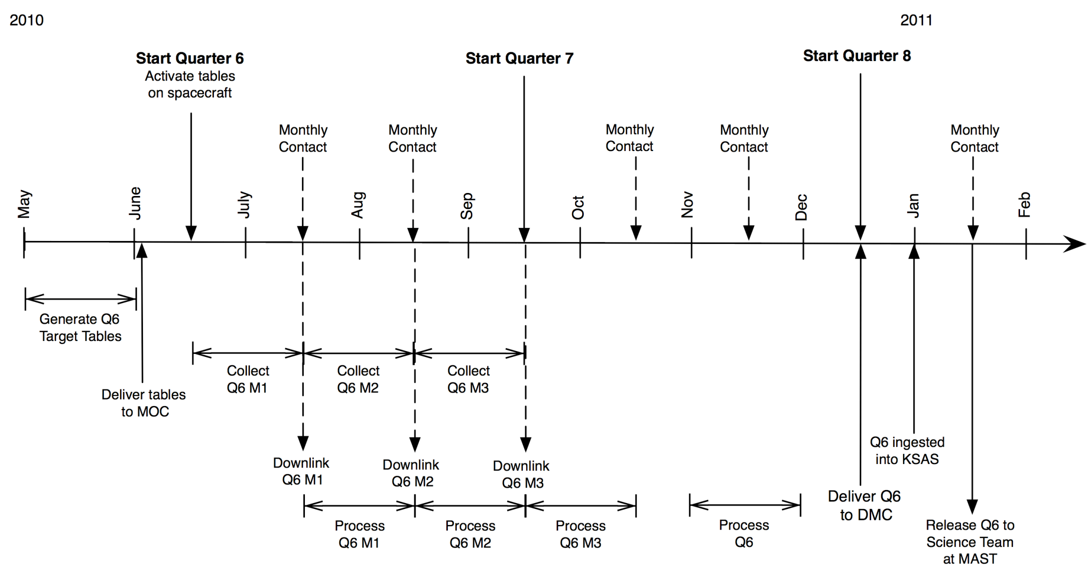

Figure 2. Routine activities for a single quarter of data

2.3 Quarterly Roll and Semi-weekly Processing

New target tables are generated each quarter and uploaded to the spacecraft because the roll of the spacecraft moves the stars to new positions on the focal plane. The LC and RP targets are observed for the full quarter, while the short cadence (SC) targets are updated monthly. After the spacecraft roll, the new target and aperture tables are activated and new science data collected. Before ending the contact with the spacecraft, a set of reference pixels are collected, downlinked, and processed through the Photometer Data Quality (PDQ) pipeline in the Flight Operations cluster at the SOC to verify the pointing of the spacecraft. Semi-weekly, X-band contact is made with the spacecraft to retrieve the reference pixel data. The reference pixels are processed through the PDQ pipeline to generate metrics on the daily spacecraft pointing and health. This data is delivered to the SO for experimentation, analysis, and reporting.

2.4 Monthly Processing

Science data are downlinked during the monthly Ka-band contacts with the spacecraft, as shown in Figure 2. The data are transferred from the Deep Space Network (DSN) to the MOC where they are unpacked then transferred to the DMC. At the DMC, the data are uncompressed and packaged into Flexible Image Transport System (FITS) format and delivered to the SOC[1]. At the SOC, the raw monthly data set is validated and ingested into the Science Processing clusters. The monthly processing of the data occurs in the *Science Processing: Monthly* Cluster (SciProcM) and takes about 10 to 15 days to complete. Monthly processing is performed to monitor instrument health, verify data completeness, understand major events, and determine parameters for quarterly processing[5]. The SOC Science Processing Pipeline (hereafter referred to as the Pipeline) calibrates the raw pixel data[6] and produces flux time series data[7,8]. The flux time series data are the photometric flux and photocenters of the ~160,000 targets. The operator uses a graphical user interface (GUI) and the SOC-developed command-line tools to configure and monitor the Pipeline. The Pipeline also generates a Photometer Performance Assessment (PPA) report on the health and performance of the spacecraft[9]. The monthly data processing is an intermediate product of the processing and is not archived. Once a full quarter's worth of science data has been collected and an initial analysis has been performed by the Science Office and the DAWG, the data are reprocessed using the Quarterly Pipeline.

2.5 Quarterly Processing

The quarterly data processing occurs in the *Science Processing: Quarterly* Cluster. The Quarterly Pipeline performs the same calibration of the pixel data and generation of the flux time series; however, it is configured slightly differently then the Monthly Pipeline to optimize the performance of the algorithms. The Quarterly Pipeline also performs the Transiting Planet Search (TPS) [10] and Data Validation (DV)[11]. See Figure 4 for the Quarterly processing workflow, which is discussed in more detail in section 3.1. The quarterly calibrated pixel data and the flux time series data is released to the Science Team and Guest Observers 120 days after downlink of the last data set, as shown in Figure 2. The reprocessing effort and export of the quarterly data set takes approximately 25 to 30 days. The DAWG reviews the data products and the Science Office Director approves it for delivery to the DMC approximately 30 days before it is due for release. The TPS and DV results from the quarterly reprocessing are ingested into the Kepler Science Analysis System (KSAS) tool for analysis, prioritization, and follow-up. The KSAS tool was developed at the Jet Propulsion Laboratory in Pasadena, CA for the Science Office. It organizes the results returned from the Pipeline and assigns Kepler Object of Interest (KOI) numbers to events identified in DV. These KOIs are then run through a merit function to prioritize the targets for ground-based follow-up observations. The SO uses the Pipeline products, KSAS, and follow-up observations to re-prioritize and clean up target lists for the upcoming quarters.

3. PROCESSES, WORKFLOWS, AND PROCEDURES

Science Operations at Ames requires close collaboration between the SO, the SOC developers, and the DAWG. The processes in place facilitate smooth transitions between data processing tasks, between development and production, and within the Kepler Ground Segment. In section 3.1, we present the Operational processes, highlighting the Target Management and Quarterly Activities. In section 3.2, we describe the issue tracking system in use at *Kepler* and how we implement it for Operations workflows. In section 3.3 we talk briefly about the procedures, defining the details and use of the tools and use cases. The processes capture the workflows and procedures into a comprehensive implementation of the science operations activities.

3.1 Processes

Kepler Operations works closely with the Science Office and other elements of the Ground Segment to support the science of the mission. Operational processes at *Kepler* rely on a combination of the procedures and the use of workflows defined in our issue-tracking system. The procedures are written with the details to accomplish a task, use cases for generic tools, acceptable deviations, and SO approval requirements. The issue-tracking system is used to log the implementation and the details specific to the task, and log any SO approvals granted.

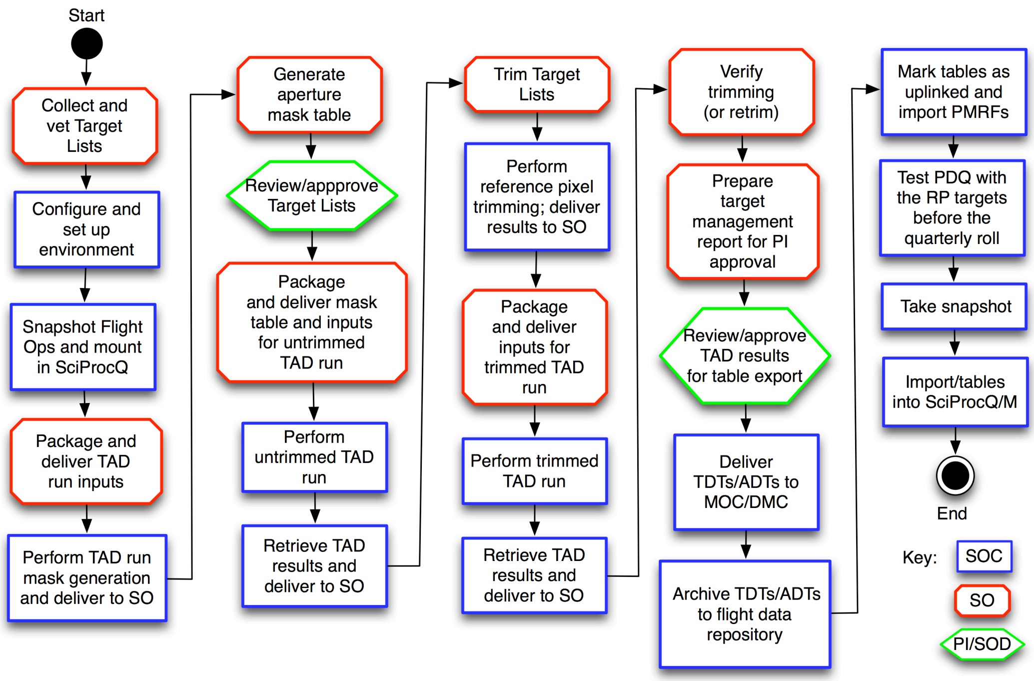

Figure 3. Target Management Activity workflow

Figure 3 demonstrates the interaction between Operations and the SO during the Target Management Activity each quarter[4]. The workflow passes between the SO Target Management scientist (octagons), the SOC operator (rectangles) and the Science Office Director (hexagons). The SO Target Management scientist and the SOC operator have separate sets of procedures to complete the Target Management Activity and several steps in the workflow occur in parallel. Operations has an overview procedure that covers the task from start to finish, calling out other procedures when necessary, such as procedures to configure a specific pipeline or export the target and aperture tables. While the SO Target Management scientist is collecting and preparing target lists, the operator is updating the FlightOps Cluster environment and preparing for the activity, such as opening a new *Target Management* ticket in the issue-tracking system. We iterate the process of target list deliveries and Pipeline processes several times to produce the final set of Target Definition Tables (TDTs) and Aperture Definition Tables (ADTs) to deliver to the spacecraft. Science Office Director (SOD) approval is required at two steps during the Target Management activity: approval of the target lists for use in the generation of the tables and approval of the final TDTs/ADTs to deliver to the MOC and DMC. After the tables have been generated, approved, and delivered, the operator performs several tasks to prepare the clusters for processing the reference pixel and science data soon to be collected on the spacecraft and making the tables available to the SO. Through the *Target Management* issue ticket, the work by each participant is tracked and the activity workflow is managed automatically by the software. At the end of the activity, the completed procedures are attached to the ticket for the records. The ticket is then resolved by the operator and closed by an external reviewer.

Figure 4 shows the workflow followed for the quarterly data processing activity that results in calibrated pixel and flux time series data for the archive at the DMC, and TPS and DV results for KSAS. A *Pipeline Processing Activity* ticket is opened for the activity and reflects the dependency on the completion of the DAWG task and the processing priorities defined by the SO and/or the operator, such as when processing multiple quarters. The ticket will also state any other dependencies on the start or completion of the quarterly processing task, such as updates to the Cluster and/or Pipeline configuration. At the start of the process, the Data Analysis Working Group (DAWG) provides Operations with updates for the configuration of the pipeline. The operator will prepare the SciProcQ cluster and configure the Pipeline for reprocessing the quarterly LC and SC data sets. Due to the manual processes involved with configuring the Pipeline, we have found it beneficial to validate the configuration of the Pipeline before launching a new instance, either by manual review from another operator or use of SOC-developed tools to validate the Pipeline configuration.

Once the data have been processed, Operations makes intermediate processing task files available and the DAWG will review and document the results for the Data Release Notes. The process of configuring the Pipeline, running the Pipeline, and DAWG review of the data is repeated for the LC and SC data processing for products to deliver to the archive, then again for the TPS and DV Pipeline for products to deliver into KSAS. An *External Data Delivery* issue ticket is created to export the products of the Pipeline and the ticket is linked to the processing ticket. With input from the DAWG, the SOD approves the *External Data Delivery* ticket before Operations ships the data to the DMC on an encrypted hard disk. The state of the data processing is communicated to the SO and DAWG members through weekly meetings with the SO, updates to the issue tickets, email, and through an internal wiki page. The wiki page contains the latest processing status of each data set, whether monthly or quarterly, and the history of the processing including details such as the cluster environment in which the data were processed and the software version used for the processing.

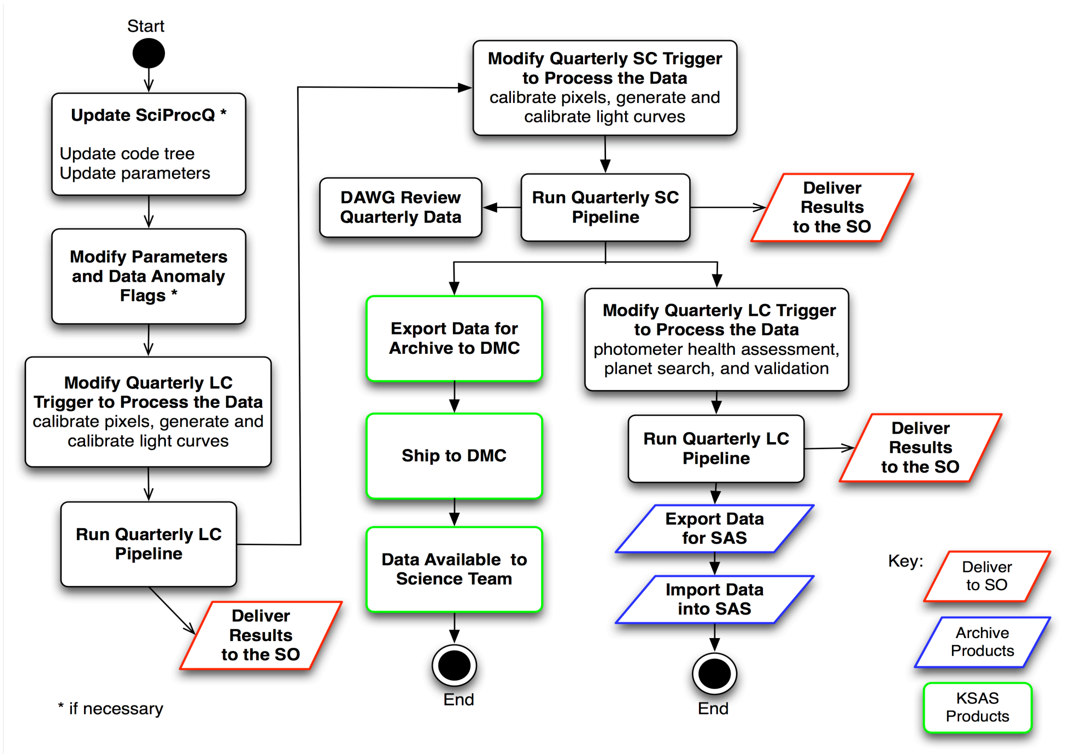

Figure 4. Quarterly Processing workflow

The SOC Pipeline Development team at Ames supports Operations by providing troubleshooting assistance, updates to the software, and training assistance. When Operations detects a software bug in the pipeline, we depend on the SOC developers to identify and correct the error. SOC development also uses the same system for issue tracking. When an error is detected, the Operator will usually file a defect ticket. This ticket is linked to the processing ticket and dependencies can be recognized. The SOC Development team will then test the software fix and patch the released software with approval from the SOC Software Review Board, which includes representatives from Operations, the SO, and the SOC Development team. The SOC developers also support the Operations through continued communication about how software changes affect the Pipeline configuration procedures. For example, the latest software release introduced the data anomaly identification flags to the Pipeline. SOC Development, Operations, and the DAWG defined the procedures by which the Operator and the DAWG identify cadences afflicted with anomalies such as thermal affects after Earth point maneuvers[8]. The operator will then update the Pipeline configuration with the data anomaly flags.

3.2 Workflows

Prior to the Commissioning phase of operations, the SOC adopted the commercial issue-tracking tool JIRA[12], which has become a valuable assistant to Operations processes. We use the issue-tracking tool to manage our operational workflows. The JIRA application supports the creation of multiple “projects” to separate the tracking of different activity types within an organization. Within each project, issue types can be created to further refine the tracking of issues within one project. At *Kepler*, we have created a project called the Kepler SOC Operations project, or KSOP. Within KSOP, we have defined 15 issue types to manage the operational workflows. The issue types range from the generic Activity/Change Request to the more specific Import FC Model issue type. Other issue types include Target Management, Test Activity, Pipeline Processing Activity, and Task. Each issue type can be defined with a unique workflow. Most of the issue types begin in the Pending OPS Review state when created. An operator reviews new KSOP tickets. Using the procedures and previous tickets of the same type as a guide, the operator creates an Implementation Plan for the activity, that lays out the high-level tasks for the SO and/or the SOC operator. The operator will then approve the activity for work and assign it to the appropriate person. Often, the KSOP tickets are used to coordinate activities that involve members of the operations staff, science office staff and SOC developers. By utilizing the Edit and Assign features, we are able to easily document the work done by each team member. The workflows support a Science Office approval loop when required by the procedure. We have also implemented the ability to mark a ticket as a "Standing KSOP." Any issue type can be marked as a Standing KSOP and be used to track routine tasks such as the semi-weekly reference pixel (RP) processing for spacecraft health monitoring. Each time the semi-weekly RPs are processed, the operator adds a comment to the KSOP with the task completion and information on where to find the results. Interested team members can receive an email for each update. Used in this manner, the standing KSOPs provide operations with both a log of the data processing and a means of notification to the interested parties. KSOP tickets are used to track the progress of a specific task, deviations, granting of SO approvals, testing, and software updates.

3.3 Procedures

SOC Operations procedures present the processes and implementation details for the SOC interactions with the DMC and MOC: table deliveries to the MOC and DMC and delivery of processed science data to the DMC. The procedures state the frequency and implementation of the deliveries. The processes explained in this paper cover a wide range of operator tasks including generation of the science target and aperture tables, semi-weekly processing of reference pixels downlinked from the spacecraft, and the quarterly processing of data for archiving at the Space Telescope Science Institute.

The Operations procedures explain the use of the tools used to perform daily to quarterly tasks and their various use cases. The procedures also note the instances when the operator must obtain SO approval before proceeding. For example, there is a single tool, and thus a single procedure, for importing the 12 Focal Plane Characteristic (FC) models into the database[13]. The procedure, called Update FC Models, defines where SO approval is necessary before importing the model into the system. The procedure states a generic usage for the command to import a model and defines the argument options to use, dependent on the model type. The procedures are written to be filled out for each use with the purpose, date, and specifics for the task. When importing an FC model, the operator will note the model being imported and the actual command used. Through the first year of operations, these procedures and processes have been refined from our experiences to better reflect operational needs and practices. Between the operational procedures and the issue tracking system, the SOC Operations team is able to use and maintain processes that support science operations at *Kepler*.

4. OPERATIONS TOOLS

The tools used to accomplish the goals of *Kepler* science operations include SOC-developed graphical user interfaces (GUIs) and command-line tools, as well as commercial products. The GUIs, discussed in section 4.1, are used for pipeline operation and configuration management. The SOC-developed command-line tools are used for many purposes ranging from data validation and ingest to Pipeline processing statistics reporting, as discussed in section 4.2. The commercial software used in Operations for cluster management and data configuration management is discussed in section 4.3. The implementation of the configuration management software for managing the products in the *flight data repository* is further detailed in section 4.4.

4.1 Graphical User Interfaces

The Science Operations Center at Kepler uses both GUI and command-line tools to support the science data processing. The Pipeline Console GUI is the operational tool used to configure and run the *Kepler* software[14]. There are three main windows of the Pipeline Console GUI: one for configuration of the pipeline and parameters, one for launching and monitoring pipelines, and one for monitoring the health and status of the cluster. The *Configuration* window of the Pipeline Console GUI is shown in Figure 5. This window manages the parameter library, module and pipeline library, data anomaly identification, and data receipt. The Pipeline Console GUI software manages the versioning of the parameters, modules, and pipelines when changed by the operator. In the *Configuration* window, the operator can add or modify parameters and pipeline definitions, view a record of the data available for processing, update data anomaly identifications, and manage user accounts. A pipeline is defined as a set of software modules that perform the desired processing tasks, along with the unit of work for each module the type of transition between modules[15]. For example, the pipeline for processing the long cadence science data consists of seven software modules, whereas the pipeline for processing the short cadence data only contains three software modules. Both the long and short cadence pipelines calibrate the pixels and then generate and calibrate the flux time series. The long cadence pipeline also produces metrics and performs photometer health analysis. The parameter library provides the flexibility to edit individual parameters manually or to import parameters from a text file. More recently added is the ability to import multiple parameters into the library via a configuration-managed XML file. The advantage of importing parameters via text and XML files is the reduced risk of transcription errors when typing in the parameter change requested by the DAWG. Once changed, parameters and pipelines are versioned and this information is associated with each pipeline instance for data accountability. A pipeline instance is a pipeline sequence that has been executed.

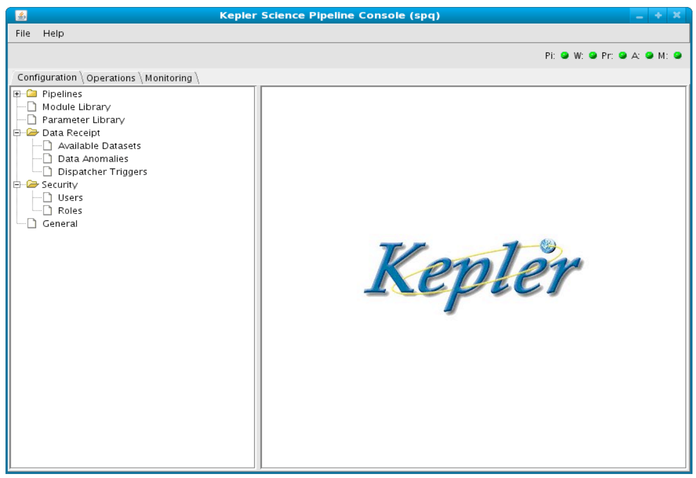

Figure 5. Pipeline Console GUI: Main login/Configuration window

The *Operations* window of the Pipeline Console GUI contains the trigger library and the pipeline instance history. Figure 6 shows the pipeline instance monitoring panel of the *Operations* window. A trigger is a pipeline configured with a specific set of parameters. It can be set to fire a new pipeline instance either manually by the operator or automatically when data is available. For example, the Photometer Data Quality (PDQ) Pipeline is configured to automatically run when new reference pixels are ingested into the Kepler Database. The Monthly and Quarterly Pipelines are configured to be manually launched by the operator. Through the trigger library pane, we can modify parameters for a trigger, validate that the trigger contains all of the necessary parameters, and save a trigger report. Once the trigger is configured for a specific data set to be processed, it can be executed and then monitored in the *Instance* pane of the *Operations* window. Once an instance of a pipeline has been launched, the parameter and software versions are tagged to that pipeline instance. In the *Instance* pane, we can monitor the progress of the pipeline instance, view the pipeline instance report and worker logs, and copy intermediate pipeline task files from individual workers. The *Monitoring* window of the pipeline GUI provides operators with the status of the cluster workers and processes, such as the Kepler Database, data receipt, and general worker processes. The *Monitoring* window also displays alerts and warnings generated during the pipeline processing. We make use of the monitoring capabilities of the Pipeline Console GUI before and during data processing. The functionalities built into the Pipeline Console GUI assists with data accountability, configuration management, monitoring of pipeline progress and cluster health, and troubleshooting errors in the pipeline. They also provide access to pipeline instance processing history, log files, and intermediate pipeline task files.

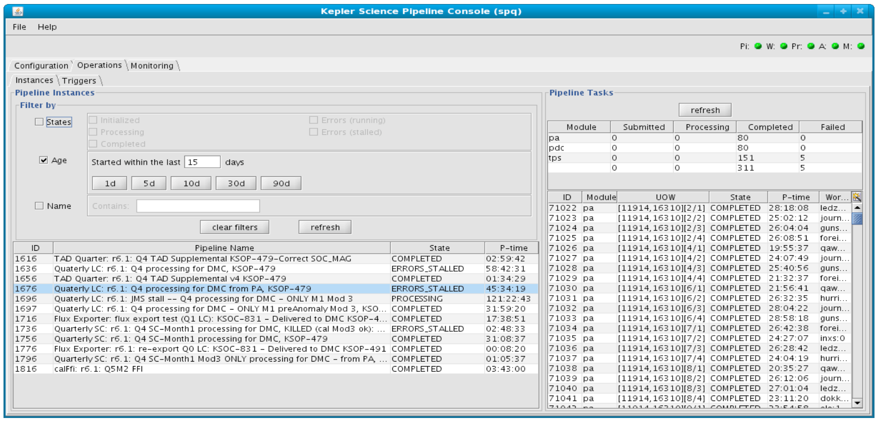

Figure 6. Pipeline Console GIU: Pipeline Instance Monitoring Window

The Target Management Console GUI is a configuration management tool used to manage and export the tables generated with the SOC Software. Figure 7 shows the main window of the tool when opened. The *Target Management* window shows the target lists imported into the database and the target list sets, which are a specified set of target lists. The tool allows for importing individual target lists, creation of target lists from a database query, definition of target lists, and export of target lists. The Target and Aperture Definition (TAD) Pipeline[4] allows for both stellar targets as identified in the Kepler Input Catalog[16] targets and custom-defined targets. Through the Target Management Console GUI, we can export the target lists with custom targets assigned a unique Kepler ID. The *Table Export* window shows the TAD, Compression, and spacecraft pointing tables available for export from the database. Through the Target Management Console GUI, the operator enters external table IDs, which will be propagated through the ground system to the spacecraft. The tables are then validated and added to the flight data repository. The SOD approves delivery of the tables to the DMC and MOC via a comment in the KSOP ticket.

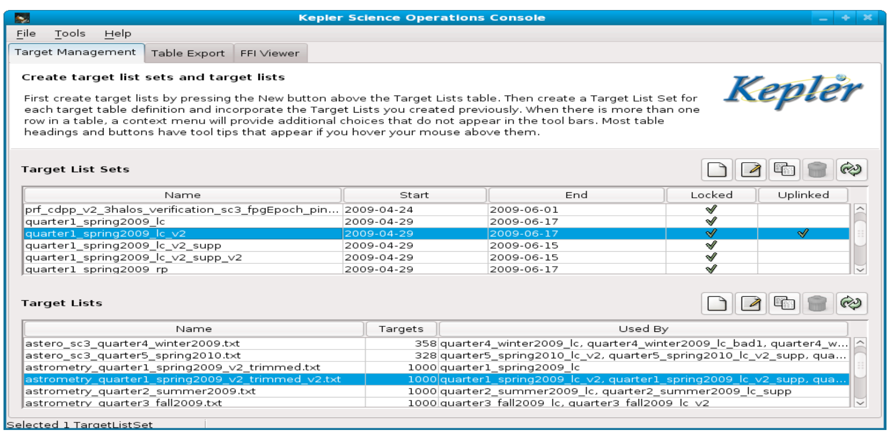

Figure 7. Target Management Console GUI

The Mission Reports (MR) GUI is a web-based tool providing the SOC and SO access to the reports generated by the Kepler Pipelines. The user accounts for MR are granted through the *Security* window of the Pipeline Console GUI. The Data Receipt and FC History links in MR retrieve the import history from the database. The operator can view the science and engineering data sets that have been imported into the Kepler Database as well as the MOC and DMC products. The FC History tool retrieves the import history of the FC models from the database. Reports available include pipeline instance reports, TAD summary reports, and PDQ reports. Spacecraft configuration maps and Compression histograms can also be viewed through MR.

4.2 Command-line Tools

Kepler operations makes use of command-line tools designed in house for several purposes such as data ingest and validation, trigger validation, and gathering pipeline instance statistics. The SOC-developed command-line tools are packaged and deployed to the processing clusters with the pipeline software. These tools are used for data accounting and validation, and pipeline configuration. There are two primary importer tools: the Focal Plane Characteristic (FC) model importer, and the data receipt (DR) tool which imports science and engineering data, target lists, and MOC and DMC products. On the command line, the operator passes to the DR tool a notification message listing the files to be imported. Each data type has a unique notification message file suffix. The DR tool first validates the data before importing it into the Kepler Database and relational database. When importing the cadence data, the DR tool validates it against the pixel mapping reference files generated at the DMC. The PMRFs were validated against the Target and Aperture tables sent to the spacecraft when they are ingested. If validation fails, the process stops and the commit to the Kepler DB or relational DB is rolled back. Each use of the DR tool is recorded in the database with the processing time and success status. Once successfully imported, the cadence data then shows as available in the Pipeline Console GUI *Data Receipt* window. The SOC-developed command-line tools provide Operations with a few tools to assist in pipeline configuration management, pipeline statistics reporting, and data accounting.

There are several scripts available to operations for validating the data and triggers. We have two tools we use to check the cadence data before we import it into the Kepler DB: the *fits report* and *validate cadence* reports. The *fits report* tool reads each cadence file and reports gaps identified in the cadence numbers and checks for inconsistencies between cadence number and MJD timestamp gaps. We can compare the gaps with what we expect based on the collection of the data on the spacecraft and the MOC report of the success of downlinking the data at the monthly contacts. The *validate cadence* tool works on a subset of the files and acts as the DR tool to verify that the files are properly formatted. The sample set is typically chosen to be files from the start of the data collection period, the middle, end, and files around anomalies that may have occurred. It is better to identify potential problems in the data before we ingest it into the database and begin processing the data. Once satisfied with the results of these reports, we ingest the data into the Kepler DB. The *rp* tool reports information on the number of reference pixels, vehicle timestamp of the data, and target table ID. This tool is most often used at the start of a quarter when we receive reference pixels from both before and after the spacecraft has rolled and the target tables have changed. For target management support, we use the *validate tad trigger* tool, which verifies that a trigger is properly configured (the planned start and end times for each Target List Set are set sequentially). The *version target list set* tool was written to aid the Target Management Activity. It clones and versions a specified target list set and re-configures the pipeline trigger to use the new target list set. The Target Management command-line tools help to reduce errors from some of the manually intensive processes in operations.

4.3 Open Source Software

We make use of several open source command-line tools to aid in Operations. The open source C3 (Cluster Command and Control)[17] software is used to configure and manage the cluster environments. We use this tool to push a code build to all workers in a cluster, as well as the operator workstations. We use it to clean up disk space on the worker machines before each Pipeline run and to monitor the status of the worker machines while the Pipeline is processing.

Subversion is an open source version control software first used in the SOC to configuration manage the SOC software[18]. Operations adopted this version control tool to configuration manage deliveries to SOC Operations, internal support products, and deliveries from SOC Operations to other elements in the *Kepler* project. We call this the flight data repository. A product under version control can be uniquely identified by a URL path and a revision number. The version control software is used to document the changes made in a file before you commit the changes to version control. The version software log features are used to track the user that made changes to a product as well as when and why the changes were made. The version control software is also used in parameter library management. Operational procedures are also maintained and managed under version control. Deliveries to SOC Operations come from several sources: the Science Office (SO), the Data Management Center (DMC), and the Mission Operations Center (MOC).

4.4 Managing the Flight Data Repository

Deliveries are organized in the flight data repository by the type of data product. Each delivery is placed in a separate subdirectory with a standard naming convention that contains the delivery date, a delivery number for that date, the submitter, and an abbreviated purpose of the delivery. The delivery often contains supplemental information about the delivery. The flight data repository deliveries are grouped into the following subdirectories: Catalogs, FC Models, Tables, Target Management, and KSAS for the SO deliveries;

Configuration Map, Leap Seconds kernel, Planetary Ephemeris, Spacecraft Clock, and Spacecraft Ephemeris for MOC products; Tables for the DMC products; and deliveries from the Guest Observer's office. When the SO has a product to deliver to Operations, they add the product to the flight data repository under version control, and notify the operators of the delivery through the appropriate KSOP ticket. In the KSOP ticket, the SO notes the Subversion URL and revision of the product. The advantage of using the version control software is the revision control feature. If a product delivered to Operations needs a minor change, the SO can simply modify the file and re-submit it to version control and update the KSOP ticket with the new revision number. Once Operations has processed the product, pipeline processing reports, logs, and supplemental products are added to the delivery directory. Routine product deliveries from the DMC and MOC are added to the flight data repository by the operator. These include products such as spacecraft ephemeris files and spacecraft configuration maps from the MOC, and the Pixel Mapping Reference Files (PMRF) from the DMC. The operator adds these files to the flight data repository and creates an "Import DMC/MOC product" KSOP with the complete file path and revision of the product.

The flight data repository is used to manage the data products delivered from SOC operations to the DMC, MOC, and Follow-up Observing Program (FOP). These products include TAD tables and Compression tables to be uploaded to the spacecraft, spacecraft pointing files sent to the MOC to re-point the spacecraft, catalogs delivered to the DMC for the archive, and follow-up observation requests sent to the FOP. Additional materials may be added to the flight data repository directory with the table for use either internally or for addition to the delivery.

The version control software is used to configuration manage the parameter library for the released *Kepler* software in use in each environment. When the DAWG identifies parameters that need to be updated, they make their modifications in a test parameter library file in the flight data repository. Once approved via KSOP, the change can be merged to the operations parameter library for one or all of the clusters before import. This helps to manage different versions of the parameters for the monthly or quarterly processing and helps to prevent transcription errors in updating parameters manually. The flight data repository is also used to manage the parameters specific to each data set. Through the Pipeline Console GUI, the operator updates, imports, and exports data-specific parameters, such as cadence range (defining the data set to be processed) and target list parameters (defining the targets to process through TPS). These parameter files are configuration managed in the flight data repository.

The flight data repository is also used to maintain supporting and reference material for operations. These include reference files, such as spreadsheets mapping the cadence and MJD range to the observation quarter, tracking spreadsheets of PDQ report reviews, and Data Release Notes for the cadence data delivered to the archive at the DMC. These references are available not only to operators, but to the SO, SOC developers, and DAWG.

5. CONCLUSION

The success of *Kepler* science operations is due to the processes and procedures in place, the use of SOC-developed tools, and the adaptation of open source software to Operations needs. The cooperation between the SO, Operations, and SOC Development has provided Operations with tools to improve validation of processes and reduce human error. The SOC-developed tools support the operational needs for easily configuring and processing the data, validation of data, and configuration management of the pipeline and data files. The issue-tracking software provides a very useful tool to maintain the Operational processes, especially those that involve multiple members of the team. As *Kepler* has worked through the first year of operations, we have made more use of the procedures, the issue-tracking tool and SOC-developed command-line tools to bring Operations to a steady state.

REFERENCES

[1] Haas, M.R., *et al.*, "Kepler science operations," *ApJL* **713(2)**, L115–L119 (2010).

[2] Middour, C., *et al*., "Kepler Science Operations Center Architecture," *Proc. SPIE* **7740**, in press (2010).

[3] McCauliff, S., *et al.*, "The Kepler DB, a database management system for arrays, sparse arrays, and binary data," *Proc. SPIE* **7740**, in press (2010).

[4] Bryson, S. T., *et al.*, "Selecting pixels for Kepler downlink," *Proc. SPIE* **7740**, in press (2010).
[5] Jenkins, J.M., *et al.*, "Overview of the Kepler science processing pipeline," *ApJL* **713(2)**, L87–L91 (2010).
[6] Quintana, E. V., *et al.*, "Pixel-level calibration in the Kepler Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[7] Twicken, J. D., *et al.*, "Photometric analysis in the Kepler Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[8] Twicken, J. D., *et al.*, "Presearch data conditioning in the Kepler Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[9] Li, J., *et al.*, "Photometer performance assessment in Kepler science data processing," *Proc. SPIE* **7740**, in press (2010).
[10] Jenkins, J., *et al.*, "Transiting planet search in the Kepler pipeline," *Proc. SPIE* **7740**, in press (2010).
[11] Wu, H., *et al.*, "Data validation in the Kepler Science Operations Center pipeline," *Proc. SPIE* **7740**, in press (2010).
[12] http://www.atlassian.com/software/jira/
[13] Allen, C., Klaus, T., and Jenkins, J., "Kepler Mission's focal plane characterization models Implementation," *Proc. SPIE* **7740**, in press (2010).
[14] Klaus, T., *et al.*, "The Kepler Science Operations Center pipeline framework," *Proc. SPIE* **7740**, in press (2010).
[15] Klaus, T., *et al.*, "The Kepler Science Operations Center pipeline framework extensions," *Proc. SPIE* **7740**, in press (2010).
[16] http://www.cfa.harvard.edu/kepler/kic/kicindex.html
[17] http://www.csm.ornl.gov/torc/C3/
[18] http://subversion.apache.org/